AFRL-AFOSR-VA-TR-2015-0404

Model-based optimal experimental design for complex physical systems

Youssef Marzouk
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

12/03/2015
Final Report

DISTRIBUTION A: Distribution approved for public release.

Air Force Research Laboratory
AF Office Of Scientific Research (AFOSR)/ RTA2
Arlington, Virginia 22203
Air Force Materiel Command

REPORT DOCUMENTATION PAGE	*Form Approved OMB No. 0704-0188*

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
30-11-2015	Final report	1 September 2012 to 31 August 2015

4. TITLE AND SUBTITLE
Final Report: Model-Based Optimal Experimental Design for Complex Physical Systems

5a. CONTRACT NUMBER

5b. GRANT NUMBER
FA9550-12-1-0420

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Marzouk, Youssef M.

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Massachusetts Institute of Technology
Cambridge, MA 02139
USA

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Jean-Luc Cambier
Program Officer, Computational Mathematics, AFOSR/RTA
875 N. Randolph St., Suite 325, Room 4104, Arlington, VA 22203

10. SPONSOR/MONITOR'S ACRONYM(S)
AFOSR/RTA

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Distribution A - Approved for Public Release

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Experimental data play an essential role in developing and refining models of physical systems. Yet experimental observations can be difficult, time-consuming, and expensive to acquire. In this context, maximizing the value of experimental observations—e.g., choosing when and where to measure, and what experimental conditions to employ—is a critical task. This project addressed open challenges in optimal experimental design (OED) for complex physical systems, taking a Bayesian decision theoretic approach. Our focus has been on general nonlinear systems and information theoretic design objectives, for which existing theory and computational tools have been inadequate. Our goal has been to develop new mathematical formulations, estimation approaches, and approximation strategies to make rigorous OED feasible for systems accessible only through computational simulation. Key results include: (1) innovations in batch optimal experimental design, in particular a new multiple importance sampling scheme that improves the efficiency of expected information gain estimation by several orders of magnitude; and (2) new dynamic programming formulations and methods for sequential optimal experimental design, supplanting previous suboptimal approaches.

15. SUBJECT TERMS
computational mathematics; optimal experimental design; uncertainty quantification; Bayesian inference; stochastic optimization; information theory; dynamic programming; estimation.

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** / **19b. TELEPHONE NUMBER** *(Include area code)*
UU	UU	UU	UU	12	Marzouk, Youssef M. / 617-253-1337

Reset

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18
Adobe Professional 7.0

INSTRUCTIONS FOR COMPLETING SF 298

1. REPORT DATE. Full publication date, including day, month, if available. Must cite at least the year and be Year 2000 compliant, e.g. 30-06-1998; xx-06-1998; xx-xx-1998.

2. REPORT TYPE. State the type of report, such as final, technical, interim, memorandum, master's thesis, progress, quarterly, research, special, group study, etc.

3. DATES COVERED. Indicate the time during which the work was performed and the report was written, e.g., Jun 1997 - Jun 1998; 1-10 Jun 1996; May - Nov 1998; Nov 1998.

4. TITLE. Enter title and subtitle with volume number and part number, if applicable. On classified documents, enter the title classification in parentheses.

5a. CONTRACT NUMBER. Enter all contract numbers as they appear in the report, e.g. F33615-86-C-5169.

5b. GRANT NUMBER. Enter all grant numbers as they appear in the report, e.g. AFOSR-82-1234.

5c. PROGRAM ELEMENT NUMBER. Enter all program element numbers as they appear in the report, e.g. 61101A.

5d. PROJECT NUMBER. Enter all project numbers as they appear in the report, e.g. 1F665702D1257; ILIR.

5e. TASK NUMBER. Enter all task numbers as they appear in the report, e.g. 05; RF0330201; T4112.

5f. WORK UNIT NUMBER. Enter all work unit numbers as they appear in the report, e.g. 001; AFAPL30480105.

6. AUTHOR(S). Enter name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. The form of entry is the last name, first name, middle initial, and additional qualifiers separated by commas, e.g. Smith, Richard, J, Jr.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES). Self-explanatory.

8. PERFORMING ORGANIZATION REPORT NUMBER. Enter all unique alphanumeric report numbers assigned by the performing organization, e.g. BRL-1234; AFWL-TR-85-4017-Vol-21-PT-2.

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES). Enter the name and address of the organization(s) financially responsible for and monitoring the work.

10. SPONSOR/MONITOR'S ACRONYM(S). Enter, if available, e.g. BRL, ARDEC, NADC.

11. SPONSOR/MONITOR'S REPORT NUMBER(S). Enter report number as assigned by the sponsoring/ monitoring agency, if available, e.g. BRL-TR-829; -215.

12. DISTRIBUTION/AVAILABILITY STATEMENT. Use agency-mandated availability statements to indicate the public availability or distribution limitations of the report. If additional limitations/ restrictions or special markings are indicated, follow agency authorization procedures, e.g. RD/FRD, PROPIN, ITAR, etc. Include copyright information.

13. SUPPLEMENTARY NOTES. Enter information not included elsewhere such as: prepared in cooperation with; translation of; report supersedes; old edition number, etc.

14. ABSTRACT. A brief (approximately 200 words) factual summary of the most significant information.

15. SUBJECT TERMS. Key words or phrases identifying major concepts in the report.

16. SECURITY CLASSIFICATION. Enter security classification in accordance with security classification regulations, e.g. U, C, S, etc. If this form contains classified information, stamp classification level on the top and bottom of this page.

17. LIMITATION OF ABSTRACT. This block must be completed to assign a distribution limitation to the abstract. Enter UU (Unclassified Unlimited) or SAR (Same as Report). An entry in this block is necessary if the abstract is to be limited.

Final Report: Model-based optimal experimental design for complex physical systems

Project duration: 1 September 2012 to 31 August 2015
Grant number FA9550-12-1-0420

Final report submitted by the Massachusetts Institute of Technology (MIT)

Chi Feng (PhD student), Xun Huan (PhD student), Youssef Marzouk (Faculty, PI)

1 Project Overview

Experimental data play an essential role in developing and refining models of physical systems. Data are used to infer model parameters, to improve the accuracy of model-based predictions, to assess the validity of models, and to improve design and decision-making under uncertainty. Yet experimental observations can be difficult, time-consuming, and expensive to acquire. In this context, maximizing the value of experimental observations—designing experiments to be optimal by some appropriate measure—is a critical task. Experimental design encompasses questions of where and when to measure, which variables to interrogate, and what experimental conditions to employ. While theory and algorithms for optimal design have been developed for many linear parameter estimation problems, rigorous and computationally tractable methods for optimal design with nonlinear simulation-based models have been sorely lacking.

This project addresses open challenges in optimal experimental design (OED) for complex physical systems, taking a Bayesian decision theoretic approach. Our focus has been on generically nonlinear systems and information theoretic design objectives, for which existing theory and computational tools have been inadequate. As described below, our goal has been to develop new **mathematical formulations**, **estimation approaches**, and **approximation strategies** to make rigorous OED feasible for systems accessible only through computational simulation. Work in this project had two major thrusts:

- *Innovations in* ***batch*** *optimal experimental design,* where all experiments are planned simultaneously before they are implemented. A key output of this thrust is a new multiple importance sampling scheme for estimating expected information gain (EIG). EIG is a central measure of the information due to an experiment, and our new estimator achieves multiple orders of magnitude smaller error (bias and variance), for a given computational effort, than previous schemes.

  Coupled with this estimator is a new formulation for **focused** experimental design, i.e., experimental design in the presence of nuisance parameters. Very often the goal of an experiment is to learn about a particular quantity of interest, yet other aspects of the system remain uncertain. Focused design maximizes the expected information gain in the marginal distribution of the parameters of interest without ignoring these other uncertainties; it can lead to very different design configurations than previous (unfocused) schemes. A natural extension of focused experimental design is the notion of experimental designs that account for model error, when model error is itself captured by a statisical inadequacy or discrepancy model.

- *New formulations and computational methods for* ***sequential*** *optimal experimental design.* Typical current practice for designing multiple experiments uses suboptimal approaches:

open-loop design that chooses all experiments simultaneously with no feedback of information, or greedy design that optimally selects the next experiment without accounting for future observations and dynamics. By contrast, sequential optimal experimental design (sOED) is free of these limitations.

We have rigorously formulated sOED as a dynamic programming (DP) problem, and developed new numerical tools to enable DP in the context of nonlinear models with continuous (and often unbounded) parameter, design, and observation spaces. Two major techniques are employed to make solution of the DP problem computationally feasible. First, the optimal policy is sought using a one-step lookahead representation combined with approximate value iteration. This approximate dynamic programming method couples backward induction and regression to construct value function approximations. It also iteratively generates trajectories via exploration and exploitation to further improve approximation accuracy in frequently visited regions of the state space. Second, transport maps are used to represent belief states, which reflect the intermediate posteriors within the sequential design process. Transport maps offer a finite-dimensional representation of these generally non-Gaussian random variables, and also enable fast approximate Bayesian inference, which must be performed millions of times under nested combinations of optimization and Monte Carlo sampling.

Collectively, this work has advanced the state of the art in optimal experimental design, yielding new computational approaches applicable to a wide range of Air Force relevant problems, ranging from object detection and inverse scattering to UAV path planning. The technical accomplishments are detailed below.

2 Technical Accomplishments

2.1 Efficient methods for focused experimental design

We examine the optimal design of experiments when the experimental goal is the inference of a *subset* of model parameters. In many scenarios, models have physical parameters and tuning parameters, but we may wish to prioritize information gain in the physical parameters over information gain in the tuning parameters.

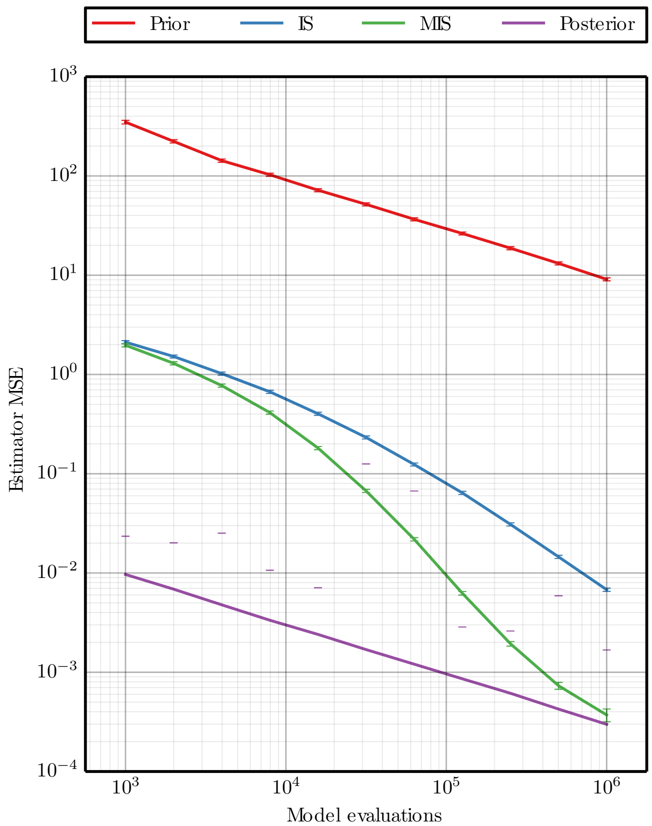
We formulate the experimental design problem in a decision-theoretic framework where the objective function is the expected information gain in only the parameters of interest. In a Bayesian setting, the information gain in the parameters of interest is represented by the difference in information carried by the prior and posterior distributions, which reflect our knowledge of the model parameters before and after observing data, respectively. Unlike existing formulations, we look at the information gain in the marginal distributions of the parameters of interest—where the influence of other so-called nuisance parameters have been integrated out—so that our objective function only reflects information gain in the parameters of interest. This allows us to exploit tradeoffs in learning between subsets of model parameters which may be overlooked if our objective were information gain in all of the model parameters.

In practice, most experimental design problems do not yield themselves to a fully analytic treatment of the expected utility. Existing approaches for estimating the expected information gain suffer from significant limitations due to computational expenses where a two order-of-magnitude gain in computational efficiency would be required even to discriminate among the enumerated designs. To this end, we have developed an efficient layered, incremental multiple importance sampling scheme for estimating the expected information gain that has the requisite orders-of-

magnitude reduction in estimator error required to make solving the exact optimal design problem tractable.

Instead of using a naive Monte Carlo estimator which draws samples from the prior distribution to estimate the posterior quantities of interest—which can be extremely inefficient when the prior and the posterior differ significantly, as is the case when data are informative—our approach incrementally approximates the posterior distribution using information from existing Monte Carlo samples that would have been discarded by the naive estimator, and remains asymptotically unbiased by using the posterior approximations indirectly as biasing distributions for unbiased importance sampling estimates. With this approach, we not only observe significant pointwise reduction in bias and variance, but we also observe a reduction in the sensitivity of the estimator bias with respect to the experimental design, which is especially important in the context of optimal design where correlations between the bias in the objective function and the design variable can result in suboptimal results.

We used our approach on two experimental design problems, a 4-dimensional linear Gaussian problem where three of the four parameters are nuisance parameters. The gain matrix has entries that are functions of the 1-dimensional design parameter that were chosen to create a clear tradeoff between designs. In Figure 1, we observe the orders-of-magnitude decrease in the estimator error, and how using the new approach allows us to find the correct optimal design at $d = 1$. The second example describes a Mössbauer spectroscopy experiment, where the goal is to choose measurement locations on the horizontal axis that result in data that are informative for inferring the parameters that describe an absorption peak. In Figure 2, we plot the posterior densities upon observing simulated data, and observe that the variance in the marginal densities for the parameters of interest are smaller when we explicity target those parameters in our focused experimental design framework.



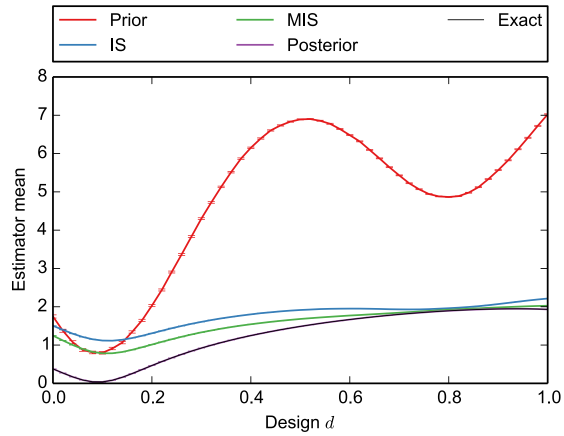


Figure 1: (Above) Estimated expected information gain for a 4D linear Gaussian model. Red is naive approach, and green is our approach. Purple is the theoretical lower bound. Note the incorrect location of the maximum when using the naive approach. (Left) Mean square error. Note the two orders-of-magnitude decrease in MSE.

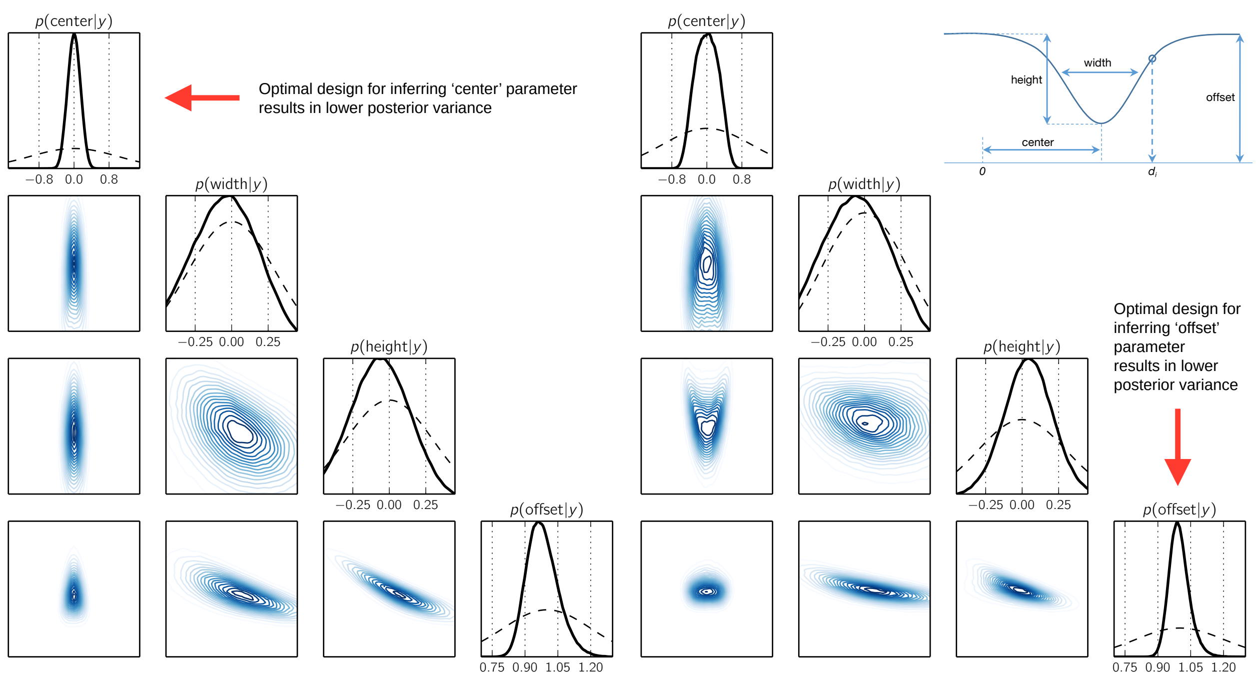


Figure 2: Our approach successfully captures the tradeoff in targeting information gain in different model parameters.

2.2 Sequential optimal experimental design

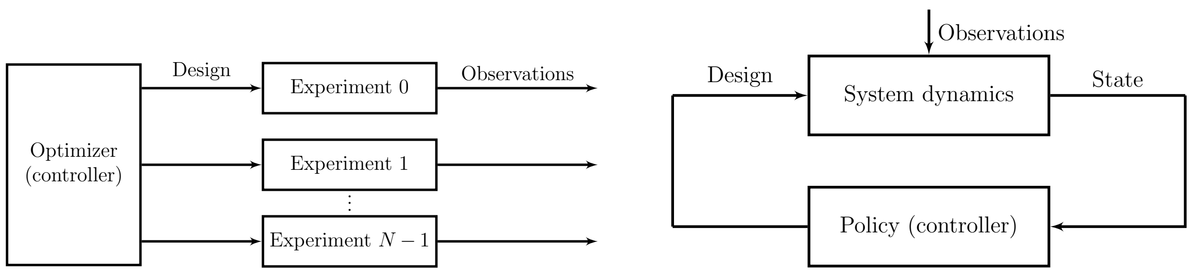
2.2.1 Formulation

Common practice for designing a sequence of experiments uses suboptimal approaches: batch design that has no feedback, or greedy (myopic) design that optimally selects only the next experiment without accounting for future effects and dynamics. The sequential *optimal* experimental design (sOED) has the advantages of

1. making use of newly acquired information during the design process to guide designs of subsequent experiments (i.e., feedback), and

2. taking into account of all future effects and dynamics.

We now seek the optimal *policy*, which consists of functions that decide what the best design is, given the updated current situation (state) (see Figure 3).



(a) Batch (open-loop) design

(b) Sequential (closed-loop) design

Figure 3: Batch design exhibits an *open-loop* behavior, where no feedback is involved, and the observations from any experiment do not affect the design of other experiments. Sequential design exhibits a *closed-loop* behavior, where feedback occurs, and the data from one experiment are used to guide the design of future experiments. "System dynamics" is the process that updates the state from one experiment to the next.

The sOED problem for N experiments involves finding the optimal policy $\pi^* \equiv \{\mu_0^*, \dots, \mu_{N-1}^*\}$, that maximizes the expected utility (reward):

$$\pi^* = \max_{\pi} \mathbb{E}_{y_0,\dots,y_{N-1}|\pi} \left[\sum_{k=0}^{N-1} g_k\left(x_k, y_k, \mu_k(x_k)\right) + g_N(x_N) \right]. \tag{1}$$

Here $d_k = \mu_k(x_k)$ is the design for the kth experiments, y_k is the observations, x_k is the state (composed of a belief state component $x_{k,b}$ that describes uncertainty, and physical state component $x_{k,p}$ that describes deterministic factors), g_k is the stage reward, and g_N is the terminal reward. The states must adhere to the system dynamics $x_{k+1} = \mathcal{F}_k(x_k, y_k, d_k)$, and the policy is subject to any design space constraints $\mu_k(x_k) = d_k \in \mathcal{D}_k$.

We focus on designing experiments to infer the model parameter θ from noisy observations y_k. To achieve this, we adopt a Bayesian perspective, and choose the belief state to be the posterior $x_{k,b} = \theta|d_0, y_0, \dots, d_{k-1}, y_{k-1}$, system dynamics to be Bayes' theorem, and terminal reward to be the Kullback-Leibler (KL) divergence from the final posterior to the initial prior—an information-measuring criterion.

Equation 1 is difficult to solve directly, but can be expressed in an equivalent form using the principle of dynamic programming (DP), that is easier to tackle:

$$\begin{aligned} J_k(x_k) &= \max_{d_k \in \mathcal{D}_k} \mathbb{E}_{y_k|x_k,d_k} \left[g_k(x_k, y_k, d_k) + J_{k+1}\left(\mathcal{F}_k(x_k, y_k, d_k)\right)\right] && (2) \\ J_N(x_N) &= g_N(x_N). && (3) \end{aligned}$$

The $J_k(x_k)$ functions are called value functions, and the optimal policy is described by the argument that maximizes the right hand side of these equations.

2.2.2 Approximate dynamic programming

Equations 2 and 3 must be solved approximately and numerically using approximate dynamic programming (ADP) techniques. We take an approach using the "one-step lookahead" representation, whose underlying idea is to construct functions $\tilde{J}_k$ that approximate J_k, for all k. Once these approximate value functions $\tilde{J}_k$ are constructed, the *approximate optimal policy* can then be extracted via one step of lookahead:

$$\mu_k(x_k) = \operatorname*{argmax}_{d_k \in \mathcal{D}_k} \mathbb{E}_{y_k|x_k,d_k} \left[g_k(x_k, y_k, d_k) + \tilde{J}_{k+1}\left(\mathcal{F}_k\left(x_k, y_k, d_k\right)\right)\right] \tag{4}$$

for $k = 0, \ldots, N-1$, and with $\tilde{J}_N(x_N) \equiv g_N(x_N)$.

We choose to numerically represent $\tilde{J}_k$ using a simple and intuitive parametric linear architecture, and construct them through the backward induction procedure:

$$\begin{aligned} \tilde{J}_k(x_k) &= r_k^\top \phi_k(x_k) \\ &= \Pi \left\{ \max_{d_k \in \mathcal{D}_k} \mathbb{E}_{y_k|x_k,d_k} \left[g_k(x_k, y_k, d_k) + \tilde{J}_{k+1}\left(\mathcal{F}_k\left(x_k, y_k, d_k\right)\right)\right] \right\} \\ &= \Pi \, \hat{J}_k(x_k). && (5) \end{aligned}$$

Here r_k is a vector of scalar coefficients, and ϕ_k are basis functions or features. The induction procedure starts at the end with $\tilde{J}_N(x_N) \equiv g_N(x_N)$ and then proceeds backwards from $k = N-1$ to $k = 1$. A suitable selection for the approximation operator Π is linear regression, which offers flexibility for generating regression points from a combination of exploration and exploitation strategies. We further developed an iterative method to improve the exploitation strategy as we gain a better understanding of the characteristics of good policies from samples generated throughout the procedure.

2.2.3 Transport maps

While ADP addresses the optimality aspect of sOED, another major difficulty remains: to numerically represent non-Gaussian, continuous random variable posteriors (i.e., belief states) that arise naturally from inference for nonlinear models. In particular, we need to use a representation that allows Bayesian inference to be performed many (millions of) times in a feasible manner under different candidate designs, observations, and priors within the ADP procedure. A suitable choice is the transport map, that is a function T that transforms a random variable z to another random variable ξ such that $\xi \stackrel{i.d.}{=} T(z)$, where $\stackrel{i.d}{=}$ denotes equality in distribution. For example, Figure 4 illustrates a log-normal random variable z mapped to a standard Gaussian random variable ξ via $\xi \stackrel{i.d.}{=} T(z) = \ln(z)$.

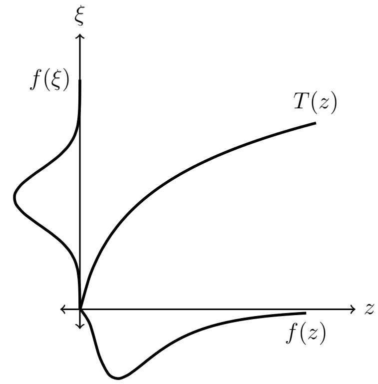


Figure 4: A log-normal random variable z can be mapped to a standard Gaussian random variable ξ via $\xi \overset{i.d.}{=} T(z) = \ln(z)$.

A special form of multivariate transport map that has a triangular structure—the Knothe-Rossenblatt (KR) map—is particularly useful for performing Bayesian inference repeatedly. For example in one experiment, with d being the design variable, y the observations, and θ the parameter to be inferred, we can construct a KR map on the joint distribution of (d, y, θ):

$$\begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{bmatrix} = \begin{bmatrix} T_d(d) \\ T_{y|d}(d,y) \\ T_{\theta|y,d}(d,y,\theta) \end{bmatrix}, \tag{6}$$

where ξ_1, ξ_2, ξ_3 are i.i.d. standard Gaussians. Because of the triangular structure of variable dependence, Bayesian inference with a particular design d^* and observations y^* simply involves conditioning (i.e., substituting) these values into the last row of the joint map, to arrive the corresponding posterior map. As a result, this inference via conditioning process can be repeated for different designs and observations at an extremely low computational cost. This concept is extended to multiple experiments, leading to a higher-dimensional joint map that allows Bayesian inference to be carried out easily for any number of experiments. The joint map is also easy to construct, as they involve solving a convex optimization problem that can be easily separated into independent sub-problems for each dimension, and requires only samples from the target distribution which is available through the aforementioned exploration and exploitation. The use of transport maps plays a crucial role in making the overall sOED method tractable.

2.2.4 Results

The sOED method is successfully demonstrated on realistic applications of optimal sensor placement. In the scenario of a chemical/biological contaminant spill, we design a sequence of locations for measuring contaminant concentrations for the purpose of inferring the contaminant source location. This sequential design problem requires a balance of planning ahead for future wind conditions, attaining high measurement signals, and reducing vehicle movement costs, which is captured withing the method developed.

In one example, we design four experiments in one-dimensional physical space. Figure 5 displays distributions of total rewards from 1000 simulated trajectories as a function of iteration for refining the exploitation measure, and the expected utility (mean) values are connected by the dashed blue line. A clear advantage of iteration is observed as the expected utility increases sharply after the first stage. Indeed, a good policy is achieved after the second iteration, and the expected utility

values of sOED are much higher than that achieved from an exploration policy (-2.0). Additionally, we have also demonstrated the advantages of sOED over batch OED and greedy design approaches (results not shown here), further supporting the near-optimality of our results.

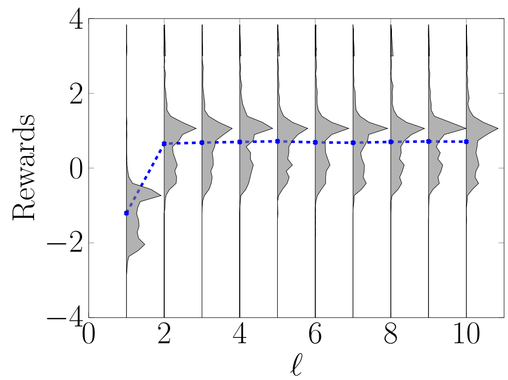


Figure 5: 1D contaminant source inversion problem: total reward distributions from 1000 simulated trajectories over increasing iterations of exploitation update ℓ. The blue dashed line connects the mean of the distributions.

In a more challenging setting, designing three experiments in two-dimensional physical space, the histograms for designs d_0, d_1, and d_2 from 1000 simulated trajectories are shown in Figure 6. Each d_k has two components, corresponding to the two physical space dimensions. The movement trend of the sensor corresponds to the expectation of future wind conditions that starts blowing to the north and northeast. The pairwise and marginal distributions from samples used to construct the joint map, and samples generated from that map are shown in Figure 7. The distributions exhibit extremely non-Gaussian, heavy-tail, and even multi-modal behavior. Nonetheless, the map is still able to capture these characteristics well, with the map-generated distributions (right) matching well with their counterparts that were used to construct the map (left).

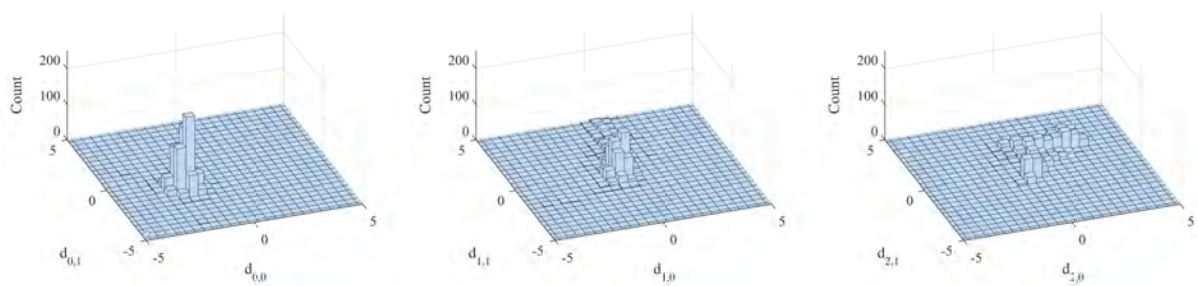


Figure 6: 2D contaminant source inversion problem: d_k histograms from 1000 simulated trajectories.

3 Personnel Supported

Two graduate research assistants (partial): X Huan, C Feng. Faculty summer support: Y Marzouk.

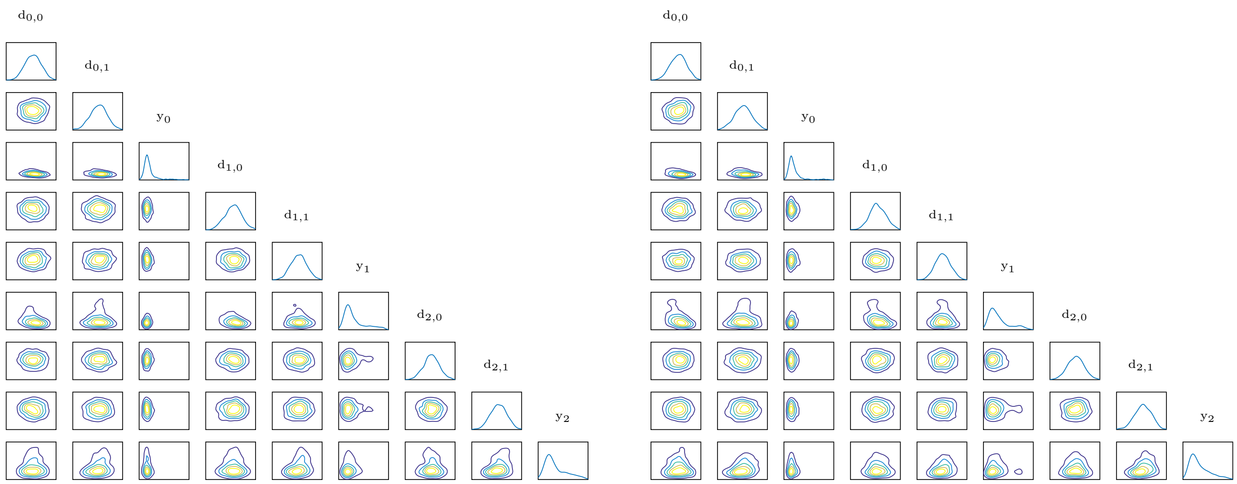


Figure 7: 2D contaminant source inversion problem: pairwise and marginal distributions from samples used to construct the exploration map (left), and from samples generated from the resulting map (right).

4 Publications and Talks

Publications

1. X Huan and YM Marzouk. "Gradient-Based Stochastic Optimization Methods in Bayesian Experimental Design." *International Journal for Uncertainty Quantification*, 4(6): 479–510, 2014.

2. X Huan and YM Marzouk. "Sequential Bayesian Optimal Experimental Design via Approximate Dynamic Programming." To be submitted to *SIAM/ASA Journal on Uncertainty Quantification*, Dec 2015.

3. C Feng, YM Marzouk. "Focused Bayesian experimental design using layered multiple importance sampling." To be submitted to *SIAM/ASA Journal on Uncertainty Quantification*, Dec 2015.

4. X Huan, M Parno, and YM Marzouk. "Adaptive Transport Maps for Sequential Bayesian Optimal Experimental Design." To be submitted to *SIAM Journal on Scientific Computing*, Jan 2016.

5. M Vohra, X Huan, and OM Knio. "Bayesian Guided Optimal Experimental Design for Reactive Multilayers." In preparation, 2015.

6. X Huan, "Numerical Approaches for Sequential Bayesian Optimal Experimental Design." PhD thesis, MIT Department fo Aeronautics and Astronautics and Center for Computational Engineering, July 2015.

7. C Feng. "Optimal Bayesian experimental design in the presence of model error." SM thesis, MIT Program in Computation for Design and Optimization, February 2015.

Talks

1. YM Marzouk, R Aggarwal, M Demkowicz, C Feng. "Bayesian optimal experimental design for materials: formulations and computational strategies." 2015 Materials Research Society

(MRS) Fall Meeting (invited speaker). Boston, MA. December 2015

2. YM Marzouk Harvard University, Department of Statistics Colloquium. Cambridge, MA. November 2015.

3. YM Marzouk Boston University, Department of Mathematics and Statistics, Statistics and Probability seminar. Boston, MA. November 2015.

4. YM Marzouk Institut Henri Poincaré. Paris, France. November 2015.

5. YM Marzouk University of Maryland, Department of Computer Science, Numerical Analysis seminar. College Park, MD. October 2015.

6. YM Marzouk Centro de Investigación en Matemáticas (CIMAT). Probability and Statistics colloquium. Guanajuato, Mexico. October 2015.

7. YM Marzouk Virginia Tech, Department of Mathematics colloquium. Blacksburg, VA. September 2015.

8. X Huan and YM Marzouk. "Optimal Sequential Experimental Design using Dynamic Programming and Transport Maps." 8th International Congress on Industrial and Applied Mathematics (invited minisymposium presentation). Beijing, China, August 2015.

9. C Feng and YM Marzouk. "Optimal experimental design in the presence of nuisance parameters and model error."13th US National Congress on Computational Mechanics (USNCCM13). San Diego, CA. July 2015.

10. X Huan and YM Marzouk. "Optimal Sequential Experimental Design using Dynamic Programming and Transport Maps." 13th U.S. National Congress on Computational Mechanics, San Diego, CA, July 2015.

11. YM Marzouk Workshop on Big Data and Predictive Computational Modeling, TU-Munich. (Invited plenary.) Munich, Germany. May 2015.

12. X Huan."Optimal Sequential Experimental Design." MIT Aerospace Computational Design Laboratory Seminar, Cambridge, MA, April 2015.

13. YM Marzouk University of California at Santa Barbara. Center for Control, Dynamical Systems, and Computation. Santa Barbara, CA. March 2015.

14. YM Marzouk King Abdullah University of Science and Technology (KAUST). Applied Mathematics and Computational Science. Thuwal, Saudi Arabia. January 2015.

15. YM Marzouk. MIT Lincoln Laboratory. Lexington, MA. August 2014.

16. X Huan and YM Marzouk. "Tractable dynamic programming for optimal sequential experimental design via approximate inference." ISBA 2014, Cancun, Mexico, July 2014.

17. C Feng and YM Marzouk. "Optimal Bayesian experimental design in the presence of model error." International Society for Bayesian Analysis (ISBA) 2014 World Meeting. Cancún, Mexico. July 2014.

18. YM Marzouk. 6th International Conference on Advanced Computational Methods in Engineering (ACOMEN). Plenary speaker. Ghent, Belgium. June 2014.

19. C Feng and YM Marzouk. "Optimal Bayesian experimental design in the presence of model error." SIAM Conference on Uncertainty Quantification (UQ14) (invited minisymposium presentation). Savannah, GA. April 2014.

20. X Huan and YM Marzouk. "Sequential Experimental Design using Dynamic Programming and Optimal Maps." SIAM Conference on Uncertainty Quantification 2014 (invited minisymposium presentation). Savannah, GA, April 2014.

21. YM Marzouk. University of Warwick, Mathematics Institute. Coventry, United Kingdom. December 2013.

22. YM Marzouk. University at Albany-SUNY. Department of Physics. Albany, NY. November 2013.

23. X Huan and YM Marzouk. "Optimal Bayesian Sequential Experimental Design using Approximate Dynamic Programming." INFORMS Annual Meeting (invited presentation). Minneapolis, MN, October 2013.

24. YM Marzouk. University of Massachusetts Dartmouth, Department of Mathematics and Center for Scientific Computing and Visualization Research. Dartmouth, MA. October 2013.

25. YM Marzouk. 38th Woudschoten Conference on Numerical Mathematics (invited keynote lectures). Zeist, The Netherlands. October 2013.

26. X Huan and YM Marzouk. "Optimal Sequential Experimental Design using Gaussian Sum Particle Filtering." 12th U.S. National Congress on Computational Mechanics, Raleigh, NC, July 2013.

27. YM Marzouk. United Technologies Research Center. East Hartford, CT. May 2013.

28. YM Marzouk. Workshop on "Numerical Methods for Uncertainty Quantification" (invited lecture). Hausdorff Center for Mathematics. Bonn, Germany. May 2013.

29. YM Marzouk. Clarkson University, Department of Mathematics. Potsdam, NY. April 2013.

30. YM Marzouk. University of Chicago, Scientific and Statistical Computing Seminar. Chicago, IL. February 2013.

31. X Huan and YM Marzouk. "Approximate Dynamic Programming for Sequential Bayesian Experimental Design." SIAM Conference on Computational Science and Engineering 2013 (invited minisymposium presentation). Boston, MA, February 2013.

32. YM Marzouk. "Surrogate modeling for uncertainty quantification." UES/Air Force Research Laboratory conference on Data Fusion for the Detection of Rare and Anomalous Events. Dayton, OH. December 2012.

33. YM Marzouk. Iowa State University, Department of Mechanical Engineering. Ames, IA. November 2012.

34. YM Marzouk. National Research Council (NRC), Workshop on "Data Collection in Support of Modeling and Simulation." Washington, DC. November 2012.

35. YM Marzouk. Air Force Research Laboratory, Materials and Manufacturing Directorate. Dayton, OH. October 2012.

36. YM Marzouk. Brown University, ICERM (Institute for Computational and Experimental Research in Mathematics) Workshop on Uncertainty Quantification. Providence, RI. October 2012.

37. YM Marzouk. University of California-Berkeley, Applied Mathematics Seminar. Berkeley, CA, October 2012.

Minisymposium organization

1. Minisymposium on "Advances in optimal experimental design for physical models." SIAM Conference on Uncertainty Quantification (UQ16). Lausanne, Switzerland. April 2016. Organizers: X. Huan, Q. Long, Y. Marzouk, R. Tempone. (12 speakers)

2. Minisymposium on "Advances in optimal experimental design." International Congress on Industrial and Applied Mathematics. Beijing, China, August 2015. Organizers: X. Huan, Q. Long, Y. Marzouk, R. Tempone. (4 speakers)

3. Minisymposium on "Advances in optimal experimental design." SIAM Conference on Uncertainty Quantification (UQ14). Savannah, GA. April 2014. Organizers: X. Huan, Y. Marzouk, L. Tenorio, G. Terejanu. (14 speakers)

Awards

1. X Huan was selected for a Natural Sciences and Engineering Research Council of Canada (NSERC) postgraduate scholarship, 2012–13.

Software releases

1. Experimental design capabilities currently in development branches of MUQ (MIT Uncertainty Quantification Library), `http://muq.mit.edu`; release in main branch expected early 2016.

1.

1. Report Type

Final Report

Primary Contact E-mail

Contact email if there is a problem with the report.

ymarz@mit.edu

Primary Contact Phone Number

Contact phone number if there is a problem with the report

6176-253-1337

Organization / Institution name

Massachusetts Institute of Technology

Grant/Contract Title

The full title of the funded effort.

Model-based optimal experimental design for complex physical systems

Grant/Contract Number

AFOSR assigned control number. It must begin with "FA9550" or "F49620" or "FA2386".

FA9550-12-1-0420

Principal Investigator Name

The full name of the principal investigator on the grant or contract.

Youssef M. Marzouk

Program Manager

The AFOSR Program Manager currently assigned to the award

Jean-Luc Cambier

Reporting Period Start Date

09/01/2012

Reporting Period End Date

08/31/2015

Abstract

Experimental data play an essential role in developing and refining models of physical systems. Yet experimental observations can be difficult, time-consuming, and expensive to acquire. In this context, maximizing the value of experimental observations---e.g., choosing when and where to measure, which variables to interrogate, and what experimental conditions to employ---is a critical task. This project addressed open challenges in optimal experimental design (OED) for complex physical systems, taking a Bayesian decision theoretic approach. Our focus has been on general nonlinear systems and information theoretic design objectives, for which existing theory and computational tools have been inadequate.

Our project has yielded new mathematical formulations, estimation approaches, and approximation strategies to make rigorous OED feasible for systems accessible only through computational simulation. Key results include: (1) innovations in batch optimal experimental design, in particular a new multiple importance sampling scheme that improves the efficiency of expected information gain estimation by several orders of magnitude; (2) focused strategies for optimal experimental design that maximize information gain only in targeted parameters of interest; and (3) new dynamic programming formulations and computational methods for sequential optimal experimental design, which vastly improve upon previous suboptimal approaches.

Distribution Statement

This is block 12 on the SF298 form.

Distribution A - Approved for Public Release

Explanation for Distribution Statement

If this is not approved for public release, please provide a short explanation. E.g., contains proprietary information.

SF298 Form

Please attach your SF298 form. A blank SF298 can be found here. Please do not password protect or secure the PDF The maximum file size for an SF298 is 50MB.

SF298_marzouk.pdf

Upload the Report Document. File must be a PDF. Please do not password protect or secure the PDF . The maximum file size for the Report Document is 50MB.

afosr_final.pdf

Upload a Report Document, if any. The maximum file size for the Report Document is 50MB.

Archival Publications (published) during reporting period:

X Huan and YM Marzouk. "Gradient-Based Stochastic Optimization Methods in Bayesian Experimental Design." International Journal for Uncertainty Quantification, 4(6): 479--510, 2014.

Changes in research objectives (if any):

None.

Change in AFOSR Program Manager, if any:

Original AFOSR Program Manager was Dr. Fariba Fahroo (Computational Mathematics Program).

In 2015, the program manager changed to Dr. Jean-Luc Cambier (Computational Mathematics Program).

Extensions granted or milestones slipped, if any:

None.

AFOSR LRIR Number

LRIR Title

Reporting Period

Laboratory Task Manager

Program Officer

Research Objectives

Technical Summary

Funding Summary by Cost Category (by FY, $K)

	Starting FY	FY+1	FY+2
Salary			
Equipment/Facilities			
Supplies			
Total			

Report Document

Report Document - Text Analysis

Report Document - Text Analysis

Appendix Documents

2. Thank You

E-mail user

Nov 30, 2015 15:21:27 Success: Email Sent to: ymarz@mit.edu